

What is claimed is:

- 1 1. A method for sending a message from a sending machine to a receiving machine,
2 comprising:
3 forming a first segment group comprising a pairing between a first source segment in
4 a memory space of a sending machine and a first target segment in a memory
5 space of a receiving machine, wherein the first source segment and the first
6 target segment comprise like number of fixed size partitions of a first partition
7 size;
8 forming a second segment group comprising a pairing between a second source
9 segment in a memory space of the sending machine and a second target
10 segment in a memory space of the receiving machine, wherein the second
11 source segment and the second target segment comprise like number of fixed
12 size partitions of a second partition size; wherein the second partition size is
13 unequal to the first partition size;
14 receiving, in the sending machine, a message to be sent;
15 selecting a segment group from the first segment group and the second segment group
16 for transferring the message, the selected segment group having a partition
17 size smaller than a length of the message;
18 copying the message into the plurality of contiguous partitions of the source segment
19 of the selected segment group; and
20 sending the content of the plurality of contiguous partitions of the source segment to
21 the receiving machine as a single message.

1 2. The method of claim 1, further comprising:
2 testing the source segment of the selected segment group to determine whether the
3 source segment is large enough to hold the message; and
4 selecting a new segment group having a partition size larger than the first selected
5 segment group, if the source segment of the first selected segment group is not
6 large enough to hold the message.

1 3. The method of claim 1, wherein the target segment is a buffer allocated in the
2 memory of the target machine and the source segment is an image of the target
3 segment memory-mapped into the memory address space of the sending machine.

1 4. The method of claim 3, wherein forming a first segment group comprising a pairing
2 between a first source segment in a memory space of a sending machine and a first
3 target segment in a memory space of a receiving machine, further comprises:
4 allocating a buffer for the target segment in memory of the receiving machine;
5 registering a descriptor of the buffer for the target segment with a remote shared
6 memory manager;
7 obtaining, at the sending machine, the descriptor of the buffer for the target segment;
8 and
9 forming an image of the target segment in the memory address space of the sending
10 machine.

1 5. The method of claim 1, further comprising:

2 registering each segment group in a segment group table and each segment in a
3 segment table.

1 6. The method of claim 5, wherein each segment comprises a directory of directory
2 entries, one for each partition in the segment, the method further comprising:
3 maintaining in an entry of the directory of directory entries an indicator of whether
4 the directory entry corresponds to a first partition of a plurality of contiguous
5 partitions storing a message and information about a next partition of the
6 plurality of contiguous partitions if the message spans multiple partitions.

1 7. The method of claim 6, wherein selecting a second paired source segment and target
2 segment having a partition size smaller than a length of the message further
3 comprises:
4 searching for a paired source segment and target segment capable of transferring the
5 message from the sending machine to the receiving machine; and
6 determining from an entry in the directory corresponding to the paired source
7 segment and target segment whether the paired source segment and target
8 segment comprise sufficient contiguous partitions to transfer the message.

1 8. The method of claim 1, further comprising:
2 forming additional segment groups comprising pairings between a source segment in
3 a memory space of the sending machine and a second target segment in a
4 memory space of the receiving machine, wherein the source segment and the

5 target segment comprise like number of fixed size partitions having a partition
6 size unequal to the first partition size and the second partition size.

1 9. The method of claim 2, wherein testing the source segment of the selected segment
2 group to determine whether the source segment is large enough to hold the message
3 comprises:
4 determining whether partition size and span factor of the segment are large enough.

1 10. A computer-readable medium carrying one or more sequences of instructions for
2 sending a message from a sending machine to a receiving machine, which
3 instructions, when executed by one or more processors, cause the one or more
4 processors to carry out the steps of:
5 forming a first segment group comprising a pairing between a first source segment in
6 a memory space of a sending machine and a first target segment in a memory
7 space of a receiving machine, wherein the first source segment and the first
8 target segment comprise like number of fixed size partitions of a first partition
9 size;
10 forming a second segment group comprising a pairing between a second source
11 segment in a memory space of the sending machine and a second target
12 segment in a memory space of the receiving machine, wherein the second
13 source segment and the second target segment comprise like number of fixed
14 size partitions of a second partition size; wherein the second partition size is
15 unequal to the first partition size;
16 receiving, in the sending machine, a message to be sent;

17 selecting a segment group from the first segment group and the second segment group
18 for transferring the message, the selected segment group having a partition
19 size smaller than a length of the message;
20 copying the message into the plurality of contiguous partitions of the source segment
21 of the selected segment group; and
22 sending the content of the plurality of contiguous partitions of the source segment to
23 the receiving machine as a single message.

1 11. The computer-readable medium of claim 10, further comprising instructions for
2 carrying out the steps of:
3 testing the source segment of the selected segment group to determine whether the
4 source segment is large enough to hold the message; and
5 selecting a new segment group having a partition size larger than the first selected
6 segment group, if the source segment of the first selected segment group is not
7 large enough to hold the message.

1 12. The computer-readable medium of claim 10, wherein the target segment is a buffer
2 allocated in the memory of the target machine and the source segment is an image of
3 the target segment memory-mapped into the memory address space of the sending
4 machine.

1 13. The computer-readable medium of claim 12, wherein the instructions for forming a
2 first segment group comprising a pairing between a first source segment in a memory

3 space of a sending machine and a first target segment in a memory space of a
4 receiving machine, further comprise instructions for carrying out the steps of:
5 allocating a buffer for the target segment in memory of the receiving machine;
6 registering a descriptor of the buffer for the target segment with a remote shared
7 memory manager;
8 obtaining, at the sending machine, the descriptor of the buffer for the target segment;
9 and
10 forming an image of the target segment in the memory address space of the sending
11 machine.

1 14. The computer-readable medium of claim 10, further comprising instructions for
2 carrying out the step of:
3 registering each segment group in a segment group table and each segment in a
4 segment table.

1 15. The computer-readable medium of claim 14, wherein each segment comprises a
2 directory of directory entries, one for each partition in the segment, and further
3 comprising instructions for carrying out the step of:
4 maintaining in an entry of the directory of directory entries an indicator of whether
5 the directory entry corresponds to a first partition of a plurality of contiguous
6 partitions storing a message and information about a next partition of the
7 plurality of contiguous partitions if the message spans multiple partitions.

1 16. The computer-readable medium of claim 15, wherein the instructions for selecting a
2 second paired source segment and target segment having a partition size smaller than
3 a length of the message further comprise instructions for carrying out the steps of:
4 searching for a paired source segment and target segment capable of transferring the
5 message from the sending machine to the receiving machine; and
6 determining from an entry in the directory corresponding to the paired source
7 segment and target segment whether the paired source segment and target
8 segment comprise sufficient contiguous partitions to transfer the message.

1 17. The computer-readable medium of claim 10, further comprising instructions for
2 carrying out the steps of:
3 forming additional segment groups comprising pairings between a source segment in
4 a memory space of the sending machine and a second target segment in a
5 memory space of the receiving machine, wherein the source segment and the
6 target segment comprise like number of fixed size partitions having a partition
7 size unequal to the first partition size and the second partition size.

1 18. The computer-readable medium of claim 11, wherein instructions for carrying out the
2 step of testing the source segment of the selected segment group to determine whether
3 the source segment is large enough to hold the message comprise instructions for
4 carrying out the step of:
5 determining whether partition size and span factor of the segment are large enough.

1 19. A system comprising:

a first programmable computer;
a second programmable computer;
a memory based interconnect for coupling the first programmable computer to the
second programmable computer by mapping one or more segments of
memory of the second programmable computer into the memory address
space of the first programmable computer;
a mechanism for forming a first segment group comprising a pairing between a first
source segment in a memory space of the first programmable computer and a
first target segment in a memory space of the second programmable computer,
wherein the first source segment and the first target segment comprise like
number of fixed size partitions of a first partition size;
a mechanism for forming a second segment group comprising a pairing between a
second source segment in a memory space of the first programmable
computer and a second target segment in a memory space of the second
programmable computer, wherein the second source segment and the second
target segment comprise like number of fixed size partitions of a second
partition size; wherein the second partition size is unequal to the first partition
size;
a mechanism for receiving, in the first programmable computer, a message to be sent;
a mechanism for selecting a segment group from the first segment group and the
second segment group for transferring the message, the selected segment
group having a partition size smaller than a length of the message;
a mechanism for copying the message into the plurality of contiguous partitions of the
source segment of the selected segment group; and

26 a mechanism for sending the content of the plurality of contiguous partitions of the
27 source segment to the second programmable computer as a single message.

1 20. An apparatus comprising:
2 a mechanism for obtaining the descriptor of a buffer allocated on a receiving
3 machine, the buffer comprising a target segment;
4 a mechanism for forming a source segment comprising an image of the target
5 segment in a memory address space;
6 a mechanism for receiving a message to be sent;
7 a mechanism for selecting a source segment for transferring the message, the
8 selection of the source segment based upon a partition size and the size of the
9 message;
10 a mechanism for copying the message into a plurality of contiguous partitions of the
11 selected source segment; and
12 a mechanism for sending the content of the plurality of contiguous partitions of the
13 source segment to a recipient machine as a single message.

1 21. An apparatus comprising:
2 a mechanism for allocating a buffer comprising a target segment in a memory;
3 a mechanism for registering a descriptor of the buffer for the target segment with a
4 remote shared memory manager; and
5 a mechanism for receiving a single message comprising content of a plurality of
6 partitions comprising a source segment image of the target segment in a
7 memory address space of a sending machine, the source segment selected

8 from among a plurality of source segments based upon a partition size and a
9 size of the message.

1 22. A method for sending a message, comprising:
2 obtaining the descriptor of a buffer allocated on a receiving machine, the buffer
3 comprising a target segment;
4 forming a source segment comprising an image of the target segment in a memory
5 address space;
6 receiving a message to be sent;
7 selecting a source segment for transferring the message, the selection of the source
8 segment based upon a partition size and the size of the message;
9 copying the message into a plurality of contiguous partitions of the selected source
10 segment; and
11 sending the content of the plurality of contiguous partitions of the source segment to a
12 recipient machine as a single message.

1 23. A method for receiving a message, comprising:
2 allocating a buffer comprising a target segment in a memory;
3 registering a descriptor of the buffer for the target segment with a remote shared
4 memory manager; and
5 receiving a single message comprising content of a plurality of partitions comprising
6 a source segment image of the target segment in a memory address space of a
7 sending machine, the source segment selected from among a plurality of
8 source segments based upon a partition size and a size of the message.

1 24. A computer-readable medium carrying one or more sequences of instructions for
2 sending a message, which instructions, when executed by one or more processors,
3 cause the one or more processors to carry out the steps of:
4 obtaining the descriptor of a buffer allocated on a receiving machine, the buffer
5 comprising a target segment;
6 forming a source segment comprising an image of the target segment in a memory
7 address space;
8 receiving a message to be sent;
9 selecting a source segment for transferring the message, the selection of the source
10 segment based upon a partition size and the size of the message;
11 copying the message into a plurality of contiguous partitions of the selected source
12 segment; and
13 sending the content of the plurality of contiguous partitions of the source segment to a
14 recipient machine as a single message.

1 25. A computer-readable medium carrying one or more sequences of instructions for
2 receiving a message, which instructions, when executed by one or more processors,
3 cause the one or more processors to carry out the steps of:
4 allocating a buffer comprising a target segment in a memory;
5 registering a descriptor of the buffer for the target segment with a remote shared
6 memory manager; and
7 receiving a single message comprising content of a plurality of partitions comprising
8 a source segment image of the target segment in a memory address space of a

9 sending machine, the source segment selected from among a plurality of
10 source segments based upon a partition size and a size of the message.